TEXT RETRIEVAL DEVICE, EFFECTIVE WORD FREQUENCY PREPARATION DEVICE, FEXT RETRIEVAL METHOD, EFFECTIVE WORD FREQUENCY PREPARATION METHOD AND RECORDING MEDIUM

Patent number: JP2000112990 (A)

Publication date: 2009-04-21

Inventor(s): OTANI NORIKO: ITO SHIRO: UEDA TAKANARI: IKEDA YUJI +

Applicant(s): CANON KK +

Classification:

- international: G06F17/30; G06F17/30; (IPC1-7); G08F17/30

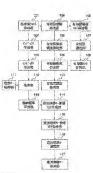
- enmnean:

Application number: JP19980300316 19981008

Priority number(s): JP19980300316 19981008

Abstract of JP 2000112990 (A)

PROBLEM TO BE SOLVED. To provide a text retrieval device and a text retrieval method which can easily re-retneval matching a retrieval object document and a retrieval result document SOLUTION. A block 104 extracts an effective word which becomes a candidate for an additional retrieval key from a retrieval object document, counts an appearance frequency in each document and prepares an effective word candidate list. Then, a block 107 refers to the effective word candidate list and prepares an effective word list. Furthermore, a block 109 deletes a word which is not an effective word from a document frequency of the effective word candidate held in a block 105 and prepares the document frequency of the effective word. A block 114 prepares an additional retrieval key candidate list from the one which corresponds to a document of the retrieval result out of the document frequency of the held effective word. A block 116 refers to the additional retrieval key candidate list and calculates an evaluation value for each additional retrieval key candidate. Then, the one whose-obtained evaluation value is high is presented as an additional retneval xey



Data supplied from the espacenet database — Worldwide

(19)日本国特許庁 (JP)

(12) 公開特許公報(A)

(II)特許出願公開番号 特開2000-112990 (P2000-112990A)

(43)公開日 平成12年4月21日(2000.4.21)

(51) Int.CL"	義別記号	F I		テーマコート*(参考)
G06F 17/30		G06F 15/401	3 1 0 A	5B075
		15/403	3408	

審査請求 未請求 謝求項の数12 FD (全 13 頁)

(21)出顯番号	特職平10-300316	(71)出職人	000001007	
			キヤノン株式会社	
(22) 出線日	平成10年10月8日(1998, 10.8)		東京都大田区下丸子3丁目30番2号	
		(72)発明者	大谷 紀子	
			東京都大田区下丸子3丁目30番2号	+1
			ノン株式会社内	
		(72)発明者	伊藤 史朗	
			東京都大田区下丸子3 1 日30番2号	44
			ノン株式会社内	
		(74)代理人	100081880	
		(弁理士 渡部 檢察	

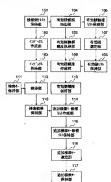
最終質に続く

(54)【発明の名称】 テキスト検索装置、有効賠額度作成装置、テキスト検索方法、及び有効賠額度作成方法並びに配 解媒体

(57) [聯約]

【課題】 検索対象文書と検索結果文書に適合した再検 素が容易に行なえるテキスト検索装置及びテキスト検索 方法を提供する。

【解決手段】 ブロック104は、検索対象交謝から追加検索キーの解析となる有効能を抽出し、各文書における出現頻繁をカウントとて有効解構制フストを作成する、そして、プロック107は、有効部候補リストを参照して有効語リストを移成する。そして、プロック10日に持ちたれた有効部域熱や文書別頻度から有効語でない語を開始して、有知語の文書別頻度を作成する、ブロック114は、保持されている有効の文書別頻度のうち、波加線を半一候補リストを参照して、各追加検索・保護リストを機がリストを参照して、各追加検索・半年(横つ資本は大きな、ブロック11日は、追加検索・年候補リストを参照して、その上加検索・半年(横つ資本は後の登りが高いるのを登加検索キーとして提示さる。



【特許請求の範囲】

【請求項1】 被検索テキストに対してインデックスを 作成するインデックス作成手段と、

前記被検索テキストから有効器候補を抽出し、各文書に おける出現頻度をカウントして有効語候補リストを作成 する有効綺樸補油出手段と、

前記有効語候補抽出手段により抽出された有効語候補の 出現輝度を文書別に保持する有効語候補額度保持手段

前記有効語撲練リストを参照して、各有効語候補の評価 10 前記検索手段により検索された検索結果を保持する検索 値を算出し、評価値の高い有効減候補を有効減として、 有効語リストを作成する有効語選定手段と、

前記有効語リストを参照し、前記有効語候補類度保持手 時に保持された有効語解補の文書別籍度から有効語でか い語を削除して、有効語の文書別網度を作成する有効語 類度作成手段と

前記インデックスにより、検索文字列に一致する文字列 を検索する検索手段と

前記有効語類度保持手段に保持された有効語の文書即翻 のから、追加検索キー候補リストを作成する追加検索キ 一解補リスト作成手段と、

前別追加検索キー候補リストを参照して、各追加検索キ 一候補の評価値を選出する追加検索キー漢字手段と、 前部追加検索キー選挙手段により求めた評価値が高いも

のを遺加検索キーントで出力する追加検索キー提示手段 とを備えたことを特徴とするテキスト検索装置。

【請求道2】 被検案テキストを保持する被検索テキス 上保持手段と、

前記被検索テキスト保持手段に保持されている被検索テ 30 キストに対して、インデックスを作成するインデックス 作或手段と:

前部インデックス作成手段で作成したインデックスを保 持するインデックス保持手段と、

前記被検索テキスト保持手段に保持されている被検索テ キストから有効誘縁縁を抽出し、各文書における出環糖 度をカウントして有効無候補リストを作成する有効語解 補抽出手段と、

前記有効語候補抽出手段により抽出された有効語候補の 出現態度を文書別に保持する有効語候補頻度保持手段

前紀有動語候補抽出手段により作成された有効語候補リ ストを保持する有効語候補リスト保持手段と、

前記有効語候補リスト保持手段に保持された有効語候補 リストを参照して、各有効語候補の評価値を算出し、評 価値の高い石効語候補を右効語として右効語リストを作 成する有効語選定手段と、

前記有効語選定手段により作成された有効語リストを保 持する有効語リスト保持手段と、

前記有効語リスト保持手段に保持された有効語リストを 50 一條補の評価値を策比する追加検索キー選定手段と。

参照し、上記有効語候補頻度保持手段に保持された有効 語候補の文書別頻度から有効語でない語を削除して、有 効請の文書別額度を作成する有効語頻度作成手段と、

前記有効語頻度作成手段により作成された有効語の文書 別類度を保持する有効語類度保持手段と、

検索を行なう文字例を保持する検索キー保持手段と、 前部インデックス保持手段に保持されているインデック

スを用いて、前記検索キー保持手段に保持されている検 素文字列に一合する文字列を検索する検索手段と

結果保持手段と

前記有効語頻度保持手段に保持された有効語の文書別類 度のうち、前記検索結果保持手段に保持された検索結果 の文書に相当するものから、追加検索キー検捕リストを 作成する追加検索キー候補リスト作成手段と、

前記職加検索キー候補リスト作成手段により作成された 追加検索キー候補リストを保持する追加検索キー候補リ スト保持手段と

前部海加極電キー候補リスト保持手段に保持された適加 度のうち、前記検索手段の検索結果の文書に相当するも 20 検索キー候補リストを参照して、各追加検索キー候補の 理係値を放出する追加検索を一部室手段と

> 前部沿加維索キー選定手段により求めた評価値が高いも のを追加検索キーとして保持する追加検索キー保持手段

追加検索キー保持手段に保持された追加検索キーを出力 する遺加権衛キ…様示手段とを鑑えたことを特徴とする テキスト検索装置。

【請求項3】 テキストから有効語候補を抽出し、各文 書における出現類度をカウントして有効語候補リストを

作贩する有効語候締抽出手段と、 前部者効能補抽出手段により抽出された者効誘能維力出 遺籍度を立憲別に保持する有効結保補額度保持手段と、

前部有効語候補リストを参照して各有効調候補の評価値 を算出し、評価値の高い有効語候補を有効類として有効 語リストを作成する有効語選定手段と、

前部有効感リストを影唱! 前記有効感解補精度保持手 段に保持された有効語候補の文書別類度から有効語でな い語を削除して、有効語の文書別額度を作成する有効語 額度作成手段とを備えたことを特徴とする有効高額度作 40 成熟源

> 【職業項4】 被検索テキストに対してインデックスを 作成するインデックス作成手段と、

> 前部インデックスにより、検索文字列に一致する文字列 を検索する検索手段と

> 請求項3記載の有効語類度作成装置によって作成された 有効熱の文書別籍度のうち、前記検索手段の検索結果の 文書に相当するものから、追加検索キー候補リストを作 成する追加検索キー候補リスト作成手段と、

前記追加検索キー候補リストを参照して、各追加検索キ

9/8/2009, EAST Version: 2.4.1.1

3 南鉄道加州検索キー選接手段により非めた評価値が高いも のを追加検索キー送上で出力する追加検索キー提示手段 とを備えたことを特徴とするテキスト検索装置。

【請求項5】 被検索テキストに対してインデックスを 作成するインデックス作成行程と、

作版するインデックス作成行権と、 前記級検索デキストから有効結候補を抽出し、各文圏に おける出現頻度をカウントして有効語候補リストを作成

する有効語候補油出行程と、 前配有効語候補抽出行程により抽出された有效語候補の 出現難度を文書別に保持する有効語候補頻度保持行程

と、 前記有効品候補リストを参照して、各有効語候補の評価 値を算出し、評価値の高い有效語候補を有効語として、 省効語リストを作成せるな効品標準行程と

前記者効語リストを案與し、前記者効語候補網度保持行 程で保持された有効語候補の文書別領度から有効語でな い語を削縮して、有効語の文書別領度を作成する有効語 頻度作成行程と、

前記インデックスにより、検索文字列に…数する文字列 を検索する検索行程と

前記有効誘頻度保持行程で保持された有效語のi文書別類 度のうち、前記模索行程の検索結果の文書に相当するも のから、追加検索キー候補リストを作成する追加検索キ 一般補リスト作成行程と、

前記追加検索キー候補リストを参照して、各追加検索キー候補の評価値を算出する追加検索キー選定行程と

前記道加校業キー選定行際により求めた評価値が高いものを選加検業キーとして出力する選加検業キー提示行程とを実行することを特徴とするテキスト検索方法。

【請求項6】 被検索テキストを保持する被検索テキス 30 ト保持行程と。

前記被検索テキスト保持行程で保持されている被検索テキストに対して、インデークスを作成するインデックス 作成行程と、

前記インデックス作成行程で作成したインデックスを保 持するインデックス保持行程と

前記被検索テキスト保持行程で保持されている被検索テ キストから有効調候補を抽出し、各文書における出現頼 度をカウントして有効請候補リストを作成する有効請候 細胞行程と

前記有効語候補抽出行程により抽出された有効語候補の 出現頻度を文書別に保持する有効語候補類度保持行程

前記有効語媒補抽出行程により作成された有効語媒補リストを保持する有効語候補リスト保持行程と、

前認有別語検補リスト保持行程で保持された有効語候補 リストを参照して、各有効語候補の評価値を算出し、評価値の高い有効語解補を対話として有効語以入トを作 値するを対法務定行程と

前記有効語選定行程により作成された有効語リストを保 第 成する追加検索キー候補リスト作成行程と、

得する有効語リスト保持行程と、

新記有効語リスト保持行程で保持された有効語リストを 参照と、上記有効語候補頻度保持行程で保持された有效 語候補の文書別頻度をから有効語でない語を理論して、有 効理のよう場別頻度を作成する有効語列度性保存程と、 地震などの表する。

前記有効語頻度作成行程により作成された有効語の文書 別頻度を保持する有効語類度保持行程と、

検索を行なう文字列を保持する検索キー保持行程と、

前記インデックス保持行戦で保持されているインテック 10 スを用いて、前記検索キー保持行程で保持されている検 楽文字列に一致する文字列を検索する検索行程と、

前記検索行程により検索された検索結果を保持する検索 結果保持行程と

的記有効論頻度保持行程で保持された有効語の文書別類 度のうち、前記極楽結果保持行程で保持された検索結果 の文書に相当するものから、追加検索キー候補リストを 作成する追加検索キー候補リスト性な行程と、

前記追加検索キー候補リスト年成行程により作成された 追加検索キー候補リストを保持する追加検索キー候補リ 30 スト保持行程と。

前點追加檢索キー候補リスト保持行程で保持された追加 検索キー候補リストを参照して、各追加検索キー候補の 評価値を施出する追加検索キー選定行程と

前記選加検索キー選定行程により求めた評価値が高いものを選加検索キーとして保持する追加検索キー保持行程と

適加検索キー保持行程で保持された適加検索キーを出力 する適加検索キー提示行程とを実行することを特徴とす るテキスト検索方法。

【請求項7】 デキストから育効語候補を抽出し、各文 書における出現領度をカウントして有効語候補リストを 作成する有効語候補抽出行程と

新記有効候補抽出行程により抽出された有効語候補の出 現様度を文書別に保持する有効語候補類度保持行程と、

約記有効語候補リストを参照して、各有効語候補の評価 値を算出し、評価値の高い有効語候補を有効語として、 有効語リストを作成する有効語繁定行程と、

制記有効論リストを参照し、前記有效論候補頻度保持行 程で保持されて有効語候補の文書別頻度から有效語でな む い語を削除して、有効話の文書別頻度を作成する有効語 頻度作成行程とを実行することを特徴とする有効語頻度 化成力が、

【請求項8】 被極素テキストに対してインデックスを 作成するインデックス作成行程と

前記インデックスにより、検索文字列に一致する文字列 を検索する検察行程と

請求用了記載の有効請額度作成方法によって作成された 有効認の文書別額度のつち、前記検索行程の検索結果の 文書に相当するものから、追加検索行子の機等結果の

前記通加検索キー業練リストを参照して、各通加検索キー 実施の評価値を算出する追加検索キー選定行程と、

的記追加検索キー選定行程により求めた評価値が高いものを追加検索キーとして出力する追加検索キー提示行程とを実行することを特徴とするテキスト検索方法。

【誘求項9】 被検索テキストに対してインデックスを 作成するインデックス作成行程と

前記被検索テキストから有効語候補を抽出し、各文器に おける出現頻度をカウントして有効語候補リストを作成 する有効語候補油出行程と、

前記有効語候補抽出行程により抽出された有効語候補の 出現頻度を文書別に保持する有効語候補頻度保持行程 と

前記者効語候補リストを参照して、各有効語候補の評価 値を算出し、評価値の高い有効語候補を有効語として、 有効語リストを作成する有効語選定行程と

前記有效語リストを参照し、前記有效語候補機度保持行程で保持された有效語候論の支書列類度から有效語でない語を開除して、有效語の支書列類度を作成する有效語報策作成行程と

前記インデックスにより、検索文字列に一致する文字列 を検索する検索行程と

前記有効需頻度保持行程で保持された有効語の文書別頻 度のうち。 勘定検索行程の検索結果の文書に相当するも のから、追加検察キー候補リストを作成する追加検案キ ・・繊維リスト性部行程と

前記道加検案キー銭徳リストを参照して、各追加検案キー鉄権の評価値を算出する追加検案キー選定行程と、

約記追加検索キー選定行程により求めた評価値が高いものを追加検索キー接定行程 かとを含むプログラムを格納したことを特徴とするコンピェーク終み取り可能な影検媒体

【請求項10】 被検索テキストを保持する被検索テキ スト保持行程と、

制記被検索テキスト保持行程で保持されている被検索テキストに対して、インデックスを作成するインデックス 作成行程と、

前記インデックス作成行程で作成したインデックスを保 持するインデックス保持行程と、

前記域検索テキスト保持行程で保持されている被検索テ 40 キストから有効語鉄種を抽出し、各文書における出現頻 度をカウントして有効語鉄種リストを作成する有効語鉄 構想出行程と、

前記有効語候補抽出行程により抽出された有効語候補の 出境無度を文書別に保持する有効語候補頻度保持行程 に

前記有効語候補抽出行程により作成された有効語候補リ ストを保持する有効語候補リスト保持行程と、

前記有効語候補リスト保持行程で保持された有效語候補 前記インデックスによ リストを参照して、各有効語候補の評価値を算出し、評 第 を検索する検索行程と

価値の高い有効語候補を有効語として有効語リストを作 成する有効語選定行程と

前記有効語選定行程により作成された有効語リストを保持する有効語リスト保持行程と、

前記有効語リスト保持行程で保持された有効語リストを 参照し、上記有効語候補類使保持行程で保持された有効 語候補の文書別類度から有效語でない語を削除して、有 効語の文書別類度をから有效語でない語を削除して、有 効語の文書別類度を作成する有効語解使作成行程と

前記有効誘頻度作或行程により作成された有効器の文書 10 別頻度を保持する有効誘頻度保持行程と、

検索を行なう文字列を保持する検索キー保持行程と、 前記インデックス保持行程で保持されているインデック スを用いて、前記検索キー保持行程で保持されている検 索文字列に一致する文字列を検索する検索行程と、 前記検索存程により検索された検索結果を保持する検索

制制(保養行程と、 結果保持行程と、 割削(有効語報度保持行程で保持された有効語の文書別報

度のうち、前記検索結果保持行程で保持された検索結果 の文書に相当するものから、追加検索キー候補リストを) 作成する追加検索キー候補リスト作成行程と、

前記追加検索キー候補リスト作成行程により作成された 追加検索キー候補リストを保持する追加検索キー候補リ スト保持行程と、

前記週加練索キー候補リスト保持行程で保持された週加 検索キー候補リストを参照して、各追加検索キー候補の 評価値を算用する追加検索キー業体行程と

前記追加検索キー選定行程により求めた評価値が高いものを追加検索キーとして保持する追加検索キー保持行程 と

適加検索キー保持行程で保持された追加検索キーを出力する追加検索キー提示行程とを含むアログラムを格納したことを特徴とするコンピュータ歳み取り可能な記録媒体。

【緯素項11】 デキストから有効額候補を抽出し、各 文書における出現頻度をカウントして有効語候補リスト を作成する有効語候補抽出行程と。

前記有効解補油出行程により抽出された有効語候補の出 現類度を文書別に保持する有効語候補観度保持行程と 前記有効語候補はストを参照して、各有効語候補の評価 質を質出し、評価値の添い有効語候補を有効語として、

有効語リストを作成する有効語選定行程と、

制起有効語リストを参照し、前記有効語候補頻度保持行 常で保持された有効語候補の文書別頻度から有效語でない語を削除して、方効語の文書別頻度を作成する石効語 頻度作成行程とを含むプログラムを格納したことを特徴 とするコンピュータ語み取り可能な記録媒体

【請求項12】 被検索テキストに対してインデックス を作成するインデックス作或行程と、

前記インデックスにより、検索文字列に一致する文字列 を検索する検索行程と 請求項11記載の記録媒体中の有效語頻度作成行程によって作成された有效語の文書別頻度のうち、商記検索行程の検索結果の文書に相当するものから、追加検索キー 鉄浦リストを作成する追加検索キー候補リスト作成行程

前記追加検索キー候補リストを奪照して、各追加検索キー候補の評価値を貸出する場所検索・選挙行程と、 前記追加検索キー選定行程により求めた評価値が高いも のを追加検索キーとして出力する追加検索キー提示行程 とを含むプログラムを指摘したことを特徴とするコンピ 10

ュータ読み取り可能な記録媒体

【発明の詳細な説明】

[0061]

【発明の概する技術分野】本発明は、検索対象文書から 検索キーに対応した文書を提示するテキスト検索装置及 びテキスト検索方法と、テキストとおける育效語の文書 別構度を作成する有效語頻度作成装置及び有效語頻度作 成方法並びに記録媒体に関する。

100021

【従来の技術】一般に、テキスト検索装置では、検索対 30 象文書から与えられた検索キーを含む文書を検索結果と して提示する。

100031

【発明が解決しようとする課題】しかしながら、上記使 来の装置では、大量に得られた検索結果を絞り込む場合 や、関連する情報をさらに集めたい場合には、適切な検 第十一を進加して検索をやり直さなくてはならないとい う問題があった。

【0004】そこで、検索キーに関連する語を追加検索 キーとして展示し、再検索を容易にすることが考えられ 第 あ、しかし、あかいの用電がおれた幕海から検索キーの 関連語を選択するのでは、辞書にない検索キーに対応で きなかったり、関連語を追加しても結果があまり変わら なかったりする。また、関連部は検索対象文書によって も異なり、検索対象文書におせて追加検索キーの候補 を決める必要がある。すなわち、再検索は、検索対象文 書と検索結果火器に適合して行う必要があるが、後来で はたれを移転に行うま法がかった。

【9005】本発明は、注記は来の問題点と議か、 検索対 東文書と検索結果文書に適合した再検索が容易に行なえ るテキスト検索装置及びテキスト検索方法を提供することを目的とする。また、テキストにおける有効語の文書 別頻度を始めかつめ頃に当成する有効顕頻度作成装置及 び有効が頻度作成方法並びに記録媒体を提供することを 目的とする。

[0006]

【課題を解決するための手段】上記目的を嘘吹するため に第1の毎明は、被検案テキストに対してインデックス を作成するインデックス年代千段と、前記波検案テキス トから春始結集解を推出し、各次素における原理頻度をラ 果保持手段に保持された検索結果外の実に相当するもの 果保持手段に保持された検索結果外の実に相当するもの

カウントして有効語候補リストを作成する有効語候補抽 出手段と、前記有効語候補抽出手段により抽出された有 効語候補の出現額度を文書測に保持する有効語候補額度 保持手段と、前記有効語候補リストを参照して、各有効 語候補の評価値を放出し、評価値の高い有効語候補を有 効語として、有効語リストを作成する有効語選定手段 と、前記有効語リストを参照し、前記有効語候補額度保 持手段に保持された有効語候構の文書別類度から有効語 でない語を削除して、有効語の文書別類度を作成する有 動語職度作成手段と、前記インデックスにより、複素文 学列に一致する文字列を練索する検索手段と、 推記有効 語頻度保持手段に保持された有効語の文書別頻度のう ち、前記検索手段の検索結果の文書に相当するものか ら 沿加検索キー候補リストを作成する追加検索キー候 補リスト作成手段と、前記追加検索キー候補リストを参 照して 各遺加検索キー候補の評価値を算出する遺加検 素キー選定手段と、前記追加検索キー選定手段によりま めた評価値が高いものを追加検索キーとして出力する追 加極索キー提売手段とを備えたものである。

【0007】第2の発明では、被検索テキストを保持す る被検索テキスト保持手段と 前記被検索テキスト保持 手簿に保持されている被練案テキストに対して インデ ックスを作成するインデックス作成手段と、前別インデ ックス作成手段で作成したインデックスを保持するイン デックス保持手段と、前記被検索テキスト保持手段に保 **終されている補検索テキストから有効誘候補を抽出**! 各立書における出現額度をカウントして有効語候補リス トを作成する有効語候補油出手段と 前記有効語候補油 出手段により抽出された有効語候補の出現頻度を文書別 に保持する有効語候補類度保持手段と、前記有効語候補 抽出手段により作成された有効議候補リストを保持する 有効誤候補リスト保持手門と、前記有効誤媒補リスト保 持手段に保持された有効諸候補リストを参照して、各有 効語候補の評価値を算出し、評価値の高い有効語候補を 有効語として有効語リストを作成する有効語選定手段 と 前記有効誘凝容手段により作成された有効誘リスト を保持する有効語リスト保持手段と、前許有効語リスト 保持手段に保持された有効誘リストを参照し、上記有効 語候補類度保持手段に保持された有効語候補の文書別類 度から有効語でない語を削除して、有効語の文書卵頻度 を作成する有効語類度作成手段と、前記有効語類度作成 手段により作成された有効感の文書別植度を保持する有 勃誘頻度保持手段と、検索を行なう文字列を保持する検 索キー保持手段と、商記インデックス保持手段に保持さ れているインデックスを用いて、前記検索キー保持手段 に保持されている検索文字列に一致する文字列を検索す る検索手段と、前記検索手段により検索された検索結果 を保持する検索結果保持手段と、前記有効語頻度保持手 段に保持された有効語の文書別類度のうち、前記検索結

から、追加検索キー候補リストを作成する追加検索キー 候補リスト作成手段と、前記適加検索キー候補リスト作 成手段により作成された追加検索キー候補リストを保持 する道加検索キー候補リスト保持手段と、前記追加検索 キー候補リスト保持手段に保持された追加検索キー候補 リストを参照して、各追加検索キー候補の評価値を算出 する追加検索キー選定手段と、前記追加検索キー選定手 段により求めた評価値が高いものを追加検索キーとして 保持する追加検索キー保持手段と、追加検索キー保持手 至手段とを備えたものである。

Q

【0008】第3の発明では、テキストから有効語候補 を抽出し、各文書における出現頻度をカウントして有効 語候補リストを作成する有効語候補抽出手段と、前記有 効候補油出手段により抽出された有効語候補の出現頻度 を文書別に保持する有効語候補類度保持手段と、前記有 効誘緊補リストを参照して各有効語候補の評価値を算出 し、評価値の高い有効語候補を有効語として有効語リス トを作成する有効顕潔定手段と、解記有効誘リストを参 照し、前記有効語候補額度保持手段に保持された有効語 候補の文書別類度から有効語でない語を削除して、有効 語の文書別額度を作成する有効機類度作成手段とを備え かものである.

【0009】第4の発明では、被検索テキストに対して インデックスを作成するインデックス作成手段と、前記 インデックスにより 検索文字列に一致する文字列を検 索する締零手的と 第3の登明の有効腫類変化成装置に よって作成された有効語の文書別類度のうち、前記検索 手段の検索結果の文書に相当するものから、追加検索キ 一候補リストを作成する追加検索キー候補リスト作成手 段と、前記追加検索キー候補リストを参照して、各追加 梅葉キー候補の評価値を無出する資加検索キー選定手段 と、前記追加検索キー漢定手段により求めた評価値が高 いものを追加検索キーとして出力する追加検索キー提示 手段とを備えたものである。

【0010】第5の発明では、被検索テキストに対して インデックスを作成するインデックス作成行程と、前配 被検索テキストから有効語候補を抽出し、各文書におけ る出環網度をカウントして有効語候補リストを作成する 有効語候補抽出行程と、前記有効語候補抽出行程により 抽出された有効誘転様の出理頻度を文書別に保持する有 **効誘接補額度保持行程と、前記有効語候補リストを参照** して、各有効語候補の評価値を奪出し、評価値の高い有 効語候補を有効語として、有効語リストを作成する有効 語選定行程と、前記有効語リストを参照し、前記有効語 候補類世保持行程で保持された有効語候補の文書別額度 から有効語でない語を削除して、有效語の文書別報度を 作成する有効語頻変作成行程と、何記インデックスによ り、検索文字列に一致する文字列を検索する検索行程

と、前記有効語頻度保持行程で保持された有効語の文書 の 【0012】第7の発明では、テキストから有効語候補

期継度のうち、前記検索行程の検索結果の文書に相当す るものから、適加検索キー候補リストを作成する適加検 素キー候補リスト作成行程と、前記追加検索キー候補リ ストを珍照して、各遺加検索キー候補の評価値を算出す る追加検索キー選定行程と、前記追加検索キー選定行程 により求めた評価値が高いものを追加検索キーとして出 力する道加検索キー標示行程とを実行するものである。 【0011】第6の発明では、波検索テキストを保持す る彼検索テキスト保持行程と、前記彼検索テキスト保持 段に保持された追加検察キ…を出力する追加検索キ…権 10 行程で保持されている被検索テキストに対して、インデ ックスを作成するインデックス作成行程と、前記インデ ックス作成行程で作成したインデックスを保持するイン デックス保持行程と、前記被検索テキスト保持行程で保 特されている物検索テキストから有効語解補を抽出し 各文書における出現網度をカウントして有効語候補リス トを作成する有効語候補油出行程と、前記有効語候補油 出行程により抽出された有効語候補の出現頻度を文書別 に保持する有効語候補額度保持行程と、歯記有効語候補 抽出行程により作成された有効議修補リストを保持する 有効語候補リスト保持行程と、前記有効語候補リスト保 持行程で保持された有効語候補リストを参照して、各有 効器候補の評価値を算出し、評価値の高い有効語候補を 有効語として有効語リストを作成する有効語選定行程 と、前記有効語漢室行程により作成された有効語リスト を保持する有効語リスト保持行程と、前記有効語リスト 保持行程で保持された有効減リストを参照し、上記有効 題候補類形保持行程で保持された有効語候補の文書別類 度から有効語でない語を削除して、有効語の文書別類度 を作成する有効語頻度作成行程と、前記有効語頻度作成 行程により作成された有効語の文書別類度を保持する有 勃開難度保持行程と、検索を行立う文字列を保持する絵 薬キー保持行程と、確記インデックス保持行程で保持さ れているインデックスを用いて、前記検索キー保持行程 で保持されている検索文字列に一致する文字列を検索す る検索行程と、前記検索行程により検索された検索結果 を保持する極端結果保持行程と 前記者勧認期度保持行 程で保持された有効語の文書例額度のうち、前影検索結 果保持行程で保持された検索結果の文書に相当するもの から、追加検索キー候補リストを作成する追加検索キー 候補リスト作成行程と、前記追加検索キー候補リスト作 成行程により作成された追加検索キー修補リストを保持 する適加検索キ…候補リスト保持行程と、前記追加検索 キー候補リスト保容行程で保持された追加検索キー候補 リストを参照して、各連加検索キー候補の評価値を算出 する遺加検索キー選定行程と、前記追加検索キー選定行 程により求めた評価値が高いものを追加検索キーとして 保持する退加検索キー保持行程と、追加検索キー保持行 程で保持された追加検索キーを出力する追加検索キー提 元行程とを実行するものである.

12

を抽出し、各文書における出現頻度をカウントして有効 議候補リストを作成する有効無候補組出行程と、再記有 効候補強出行程により抽出をれた有効無候補物の出現頻度 を文書別に保持する有効無候補和使保持行程と、前記有 効品候補リストを参照して、各有効品候補の評価値を築 出し、評価位の高い有効品候補を有効過として、有効品 リストを作成する有効部型社行程と、前記有効品リストを 参照し、語な力効論候補利度持行程で発きれた有 効語候補の文書別頻度から本分が語でない語を利除して、 有効品の文書別頻度から本方効語研究性を を指行るものである。

【6013】第8の売即では、被検索テキストに対して インデックスを性咳するインデックス性低行程と、前記 インデックスにより、検索文字列に一致する文字列を検 索する検索行程と、上記等での発明の育効誘卵度作成方 法によって作成された有効態の文書別頻度のうち、前記 検索行程の機能規定交通に担当するものから、追加検 素キー候補リストを作成する追加検索キー維施リスト作 成行程と、前記追加検索キー展施と多りましたそ が高いものを通加検索キー選定行程により求めた評価値 が高いものを通加検索キー選定行程により求めた評価値 が高いものを通加検索キーとして出力する通加検索キー 提示行程とを発行するものでちる。

【0014】第9の発明は、破検索テキストに対してイ

ンデックスを作成するインデックス作成行程と、前記被

強索テキストから有効振解補を抽出! 各文書における 出現駆的をカウントして有効誘爆補リストを作成する有 効語候補拘出行程と、前記有効語候補輸出行程により抽 出された有効語候補の出現頻度を文書明に保持する有効 語候補物度保持行程と、前記有効語候補リストを参照し て 各有効消除額の評価値を算出! 評価値の高い有効 語候補を有効訊として、有効語リストを作成する有効語 選定行程と、前記有効語リストを参照し、前記有効語儀 補類度保持行程で保持された有効語候補の文書別類度か ら有効語でない語を測除して、有効語の文書別頻度を作 成する有効誘頻能作成行程と、前記インデックスによ り 検索文字列に 費する文字例を検索する検索行程 と、前記有効誘線選保特行程で保持された有効語の文書 別類度のうち、前記検索行程の検索結果の文書に相当す るものから、追加検索キー候補リストを作成する追加検 む 零キー候補リスト作曲行程と 前別追加検索キー保練り ストを参照して、各追加検索キー候補の影価値を算出す る適加検索キー選定行程と、商売追加検案キー選定行程 により求めた評価値が高いものを追加検索キーとして出 力する追加検索キー提示行程とを含むプログラムを格納

【60日5】第10分発明は、被検索テモストを保持す る放検電テモスト保持行程と、耐定接検電テモストとは持 勿能解除加立て業別頻度から有効器でない 有効器の文章別頻度から有効器でない 有効器の文章別頻度から有効器でない 有効器の文章別頻度から表別 を申収するインデックス作取行程と、耐湿インデ っ

したものである。

ックス作成行程で作成したインデックスを保持するイン デックス保持行程と、前記飯検索テキスト保持行程で保 持されている被検索テキストから有効語候補を抽出し、 各文書における出現額度をカウントして有効語候補リス トを作成する有効語候補抽出行程と、前記有効語候補抽 出行程により抽出された有効は候補の出現額度を文書別 に保持する有効語候補類度保持行程と、前記有效語候補 抽出行程により作成された有効語候補リストを保持する 有効語候補リスト保持行程と、前記有効語候補リスト保 持行程で保持された有効語候補リストを参照して、各有 勃浪候補の評価値を算出し、評価値の高い有効試候補を 有効語として有効語リストを作成する有効語選定行程 と、前記有効語選定行程により作成された有効語リスト を保持する看効語リスト保持行程と 確認者効語リスト 保持行程で保持された有効語リストを参照し、上記有効 語候補額度保持行程で保持された有効語候補の文書別籍 度から有効舗でない舗を削除して、有効舗の支書別類度 を作成する有効語頻度作成行程と、前記有効語頻度作成 行程により作成された有効語の文書別類度を保持する有 効誘駆度保持行程と、検索を行なう文字列を保持する検 素キー保持行程と、前記インデックス保持行程で保持さ れているインデックスを用いて、前記検索キー保持行程 で保持されている物素文字列に一致する文字列を検索す る検索行程と、前記検索行程により検索された検索結果 全保持する検索結果保持行程と、前記有効語幅度保持行 程で保持された有効熱の文書問題度のうち、前記検索は 果保持行程で保持された極零結果のウ素に相当するもの から、裏加検索キー候補リストを作成する過加検索キー 候補リスト作成行程と、前記退加検索キー候補リスト作 成行程により作成された追加検索キー候補リストを保持 する密加極歌キー候補リスト保持行程と 前記追加検索 キー候補リスト保持行程で保持された追加機索キー候補 リストを参照して、各追加検索キー候補の評価値を算出 する追加検索キー選定行程と、前記追加検索キー選定行 程により求めた評価値が高いものを追加検索キーとして 保持する品加検索キー保持行程と 追加検索キー保持行 程で保持された適加検索キーを出力する適加検索キー様 示行程とを含むプログラムを格納したものである。

【0016】第11の範別は、デキストから有効需候補 を抽出し、各次常における出現頻度をカウントして有効 部候補のストを作成する有効高候補油出行程と、前記有 効候補油は行程により抽出された有効高候補の出現頻度 を文書別に保持する有効高候補規度保持行程と、前記有 が高候補リストを実践して、各有効高候稀の評価値を集 出し、評価値の高い有効高度接種を有効語として、有物語 リストを作成する有効高低深程と、前記有が高りスト 参照し、前記有効高候端補度保持行程で保持された 効態候補の文書別頻度から表別でない結を削壊して、 有効語の文書別頻度から有効高切でない結を削壊して、 有効語の文書別頻度から有効高切でない結を削壊して、 13

【0017】第12の発眼は、被検索テキストに対して インデックスを作成するインデックス作成行程と、前記 インデックスにより、検索文字列に…致する文字列を検 索する検索行程と、上記第11の発明の記録媒体中の有 **効語頻度作成行程によって作成された有効語の文書別類** 度のうち、前記検索行程の検索結果の文書に相当するも のから、追加検索キー候補リストを作成する追加検案キ 一候補リスト作成行程と、前記追加検索キー候補リスト を参照して、各追加検索キー候補の評価値を算出する追 加検索キー選定行程と、前記巡加検索キー選定行程によ 10 り求めた評価値が高いものを追加検索キーとして出力す る追加検索キー提示行程とを含むプログラムを格納した ものである。

[0018]

【発明の実練の形態】以下、図面を参照して本発明の実 籐の形態を説明する、

【0019】 [第1実練形骸] 図1は、本路明の第1実 維形機のテキスト検索禁制の機能機成を示すプロック図 である。

【0020】開閉において、101は、披練物テキスト 20 を保持する紡権電デキスト保持部である。10つは、被 検索テキスト保持部1の1に保持されている紡績索テキ ストに対して一般素を高速に行なっためのインデックス を作成するインデックス作成部である。103は イン デックス作成部102で作成したインデックスを保持す るインデックス保持部である。104は、紡権電テキス ト保持部101に保持されている被検索テキストから有 効語候補を抽出し、各文書における出現糊度をカウント して有効語候補リストを作成する有効語候補抽出部であ る。有効語候補抽出部104は、被検索デキストを形態 30 素解析! その結果 名詞と解析された文字列を育動語 候補とする。105は、有効部候補抽出部104により 抽出された有効議候補の出現難度を文書別に保持する有 **勤語候補和度保持部である。106は、有効語候補抽出** 部104により作成された有効語候補リストを保持する 有効誤媒種リスト架棒部である。有効誘媒種リストは 有効原候補ごとに各文書における出環額度を保持してい る。有効語候補リストの例を図えに定す。

【0021】107は有効語選定部であり、有効語候補 昭して各有効語候補の評価値を放出!. 評価値の高い有 効調整補を有効額として有効語リストを作成する。有効 語リストの何を図3に示す。評価値は、各有効語候様の 重要度を示す値である。各文書を特徴付ける語が重要な ので、検索対象文書内における各語の局在液を評価値に 反映させるべきであるが 局有度そのものを評価値にす ると、ある文書にたまたま一度出現した語の評価値が高 くなってしまう。そこで、ある程度の出現頻度があり、 かつ局在度の高い語を有効語とするために、出現頻度と 局在度の積を評価値とする。

【0022】有効語候補買すを含む文書数を料とする と、Wiの評価値E(Wi)は次の式により求められ 2

[0023]

(数1)

$$E(W_i) = M \cdot (1 + \sum_{j=1}^M P_{i,j} \log_M P_{i,j})$$

$P_{i,i} = \frac{\dot{\chi} \hat{\mathbf{S}} D_i \mathbf{P} i \pi i \mathbf{D} t \mathbf{D} \mathbf{V}_i$ の出現頻度 文書 D,中の有効語候補数

108は、有効諸選定器107により作成された有効語 リストを保持する有効語リスト保持網である。109は 有効語頻度作成部であり、有効語リスト保持部108に 保持された有効語リストを参照し、有効語候補頻度保持 部10円に保持された有効誘蜉補の文書別籍度から有効 語でない語を削除して、有効語の文書別頻度を作成す 8. .

【0024】110は、有効誘頻度作成部109により 作业された有効頭の文書別網度を保持する有効誘頻度保 特部である。111位 練案を行かう文字刷を保持する 検索キー保持部である。112は検索部であり、インデ ックス保持部103に保持されているインデックスを押 いて、検索キー保持部111に保持されている検索文字 列に一致する文字列を検索する、113は、検索部11 2により検索された検索結果を保持する検索結果保持部 である。

【0025】11日は適加検索キー候補リスト作成部で あり、有効語頻度保持部110に保持された有効語の文 書別難度のうち。 検索結準保持部113に保持された検 薬結果の文書に相当するものから、 追加検索キー餘補り ストを作成する。遺加検索キー候補リストは、有効調候 補リストと同様の構造で表現できる。115は、追加検 索キー候補リスト作成部114により作成された追加検 索キー解析リストを保持する湯加物索キー修練リスト保 特部である.

【0026】116は追加検索キー選定部であり、追加 検索キー候補リスト保持部115に保持された追加検索 リスト保持部106に保持された有効諸候補リストを参 む キー候補リストを参照して、各進加検索キー候補の評価 値を算出する、評価値は 各追加検索キー保護の再検索 における有効度を示す値である。検索結果を絞り込むた めに再検索を行なうには、検索結果文書において局在度 の高い語が有用と考えられるので、退加検索キー候補W を含む検索結果文書数を別としたとき、Wiの評価値目 (Wi) は次の式により求められる。

[0027]

【数2】

$$E(W_i) = M \cdot (1 + \sum_{i=1}^{M} P_{i,j} \log_M P_{i,j})$$

文書 D;中における W,の出現頻度 文書 D.中の追加検索キー候補数

117は、追加検索キー選定部116により求めた評価 値が高いものを追加検索キーとして保持する追加検索キ …保持部である。

【0028】四4は、水実施形機に係るテキスト検索装 置のハードウェア構成を示す図である。

【0029】 岡図において、201は、後述する制御手 順を実現するプログラムを保持するROMである。20 2はRAMであり、被検索テキスト保持部101。有効 語鏡補類度保持部105、有効語候補リスト保持部10 6、有効語リスト保持部108、検索キー保持部11 1、検索結果保持部113、追加検索キー候補リスト保 将郷115 及び追加検索キー保持郷117と 上記プ ログラムの動作に必要な記憶領域とを提供する、

【0030】203は、ROM201に保持されている プログラムに従って処理を行たう中央処理装置である。 204にディスク装置であり、インデックス保持部10 3、有効語頻度保持部110を実現する。205はバス であり、上記の各構成を接続し、各構成間におけるデー タの母母を可能とする。

【0031】次に、本誌器の動作を説明する。

【0032】本実練形線の原理は、インデックス作成フ ェーズと検索フェーズとに大きく分かれる。以下、図5 のフローチャートを参照して、本実施形態のテキスト検 36 妻装置におけるインデックス作成フェーズの手順を説明 18

【0033】ステップS301では、カウンタ主の初期 化を行なう。カウンタ主は、処理の対象となっている被 検索テキストの文書番号を示すもので、これをりに初期 化する。そしてステップ8302に移る。ステップ83 0.2では、カウンタトが教検索テキストの文書数に達し たか否かを判定して、達した場合はステツブS308に 移る。達していない場合は、ステップS303に移る。

第テキストのインデックスを作曲する。 そしてステップ S304に移る。ステップS304では、文書番号:の 被検索テキストから、有効語候補となる文字列とその文 書内出現頻度を抽出する。そしてステップS305に移 る。ステップS305では、文書番号iの被検索テキス トに出現する有効語候補とその類度を保存する。そして ステップS306に移る。ステップS306では、文書 番号主の被検索テキストの有効語候補に関する情報によ り、有効語候補リストを更新する。そしてステップS3 07に移る。

【0035】ステップ3307では、カウンタ主の値を 1 増やす。そしてステップS302に関る。ステップS 308では、各有効語候補の評価値を算出する、そして ステップS309に移る。

【0036】ステップS309では、ステップS308 で求めた評価値の高いもの、隣えば、評価値が所定値以 上のものを有効語とし、有効語リストを作成する。そし てステップS310に移る。ステップS310では、ス テップド305で文書ごとに保存した有効語候補とその 10 頻度から、有効語でないものを削除し、有効語とその類 **寝として保存する。そしてインデックス作成フェーズを** 終了する。

【0037】次に、図ものフローチャートを参照して、 本実施形態における検索フェーズの概要を発明する。 【0038】ステップS401では、検索キーを含む文 書を検索する。そしてステップS402に移る。ステッ アS402では、カウンタ主の初期化を行なう。カウン タiは、ステップS401で得られた検索結果の結果番 考を示すもので、これを0に初期化する。そしてステッ 20 プS403に移る。ステップS403では、カウンタ1 が検索結果数に達したか否かを判定して、達した場合は ステップS406に移る。達していたい場合はステップ S404に移る。

【0039】ステップS404では、i番目の検索結果 文書の有効語とその出現頻度を追加検索キー候補リスト は追加する。そしてステップS405に移る。ステップ S305では カウンタ主の値を1増やす。そしてステ ップS403に戻る。ステップS406では、各追加検 索キー候補の評価値を算出する、そしてステップS40 7に移る。ステップS407では、ステップS406で 主のた評価値の高いものを追加検索キーとする。そして 検索フェーズを終了する。

【0040】以上に述べたように、検索対象文書から選 加検索キーの候領となる有効語を抽出し、検索結果文書 における有効語の出現状況を反映した追加検索キーを提 示することで 検索対象文書と検索結果文書に適合した 再検索が容易に行なえるテキスト検索装置を実現するこ とができる。

【0041】「第2実施形験] 上記第1実権形態におい 【0034】ステップS303では、文書番号iの被検 む では、インデックス作成フェーズと検索フェーズを詞一 装置で行なう場合について説明したが、これに限定され るものではない。例えば、インデックス作成フェースだ けを行なう装置であってもよい。図ではこの場合を示す 本を明の第2字論形構の基本的な機能構成を示すプロッ ク図である。

> 【0042】例7において、701は、被検索テキスト を保持する被検索テキスト保持部である。702は、被 検索テキスト保持部701に保持されている被検索テキ ストに対して、検索を高速に行なっためのインデックス 50 を作成するインデックス作成部である。703は、イン

デッタス作成部702で作成したインデックスを保持す るインデックス保持部である。704は有効語候補抽出 部であり、被検索テキスト保持部701に保持されてい る被検索テキストから有効語候補を抽出し、各文書にお ける出現頻度をカウントして有効語候補リストを作成す る 7 0 5(t. 有効新解補額度保持部であり、有効系候 補抽出部704により抽出された有効語候補の出現額度 を文書別に保持する。

【0043】706は有効顕像補リスト保持部であり 有効誤候補排出部704により作成された有効誘候結り 10 インデックスと有効試練度が、通常回線或いは可搬影縁 ストを保持する。707は有効減避定部で売り、有効額 候補リスト保持部706に保持された有効語候補リスト を参照して 各有効議候補の評価値を貸出し 評価値の 高い省独語候補を有効語として、有効語リストを作成す

【0044】708は有効語リスト保持部であり、有効 語選定部707により作成された有効語リストを保持す る。709は有効語頻度作成部であり、有効語リスト保 将郷708に保持された有効離リストを参照し、有効語 語候補種度保持部70万に保持された有効語候補の文書 20 別頻度から有効語でない語を削除して、有効語の文書別 制度を作成する。710日右が議額度保持部であり、右 効誘駆停作成部709により作成された有効語の文書期 物度を保持する。

【0045】インデックス保持部703に保持されたイ ンデックスと有効調整保持部710に保持された有効 謎の文書別顧序は 通信問線を通したり 可機型経媒体 によって他の装置に移されて検索が行なわれる

【0046】「第3実練形態」上述のようにして提供さ れたインデックスと有効語の文書別類度を担いて検索を 36 行なっ装置であっても上い。 翔らはこの場合を元す本業 明の第3実験形態の基本的な機能構成を示すプロック図 である。

【0047】図8において、801は、図7に示す装置 で作成されたインデックスを保持するインデックス保持 部である。80.2は 検索を行かう文字刷を保持する検 素キー保持部である、803は検索部であり、インデッ クス保持総801に保持されているインデックスを用い て、検索キー保持部802に保持されている検索文字列 に一致する文字列を検索する。

【0048】804は 検索部803により検索された 検索結果を保持する検索結果保持部である 805は、 図7に示す装置で作成された有効層の文書別幅度を保持 する有効語頻度保持部である。806は適加検索キー候 補リスト作成部であり、有効語類度保持部805に保持 された有効語の文書明頻度のうち、検索結果保持部80 4に保持された検索結果の文書に相当するものから、追 加検索キー候補リストを作成する。

【0049】807は、追加検索キー候補リスト作成部 806により作成された資加検索キー候補リストを保持 50 限定されるものではなく ステップ 5306の移にイン

する油油検索キー候補リスト保持窓である。808は退 加検索キー選定部であり、遅加検索キー候補リスト保持 部807に保持された追加検索キー候補リストを参照し て、各追加検索キー候補の評価値を算出する。809は 追加検索キー保持部であり、追加検索キー選定部808 により求めた評価値が高いものを追加検索キーとして保 握する。

【0050】インデックス保持部801と有効語類度保 持部805には、上記第1実維形態の手順で作成された 媒体等を介して提供され、保持される。

【0051】なお、本発明は選示の実施形態に限定され ず様々の変形が可能である。 例えば その変形例として は次のようなものがある。

【0052】(1)上記第1実練形態では、検索結果を 絞り込むために再柳索を行なうものとして、遺加検索キ 一院補の評価値を求めるよう説明したが、これに限定さ れるものでなく、以下の式により評価値を求め、検索条 件を綴れて検索結果を順やすための再編案を行むう場合 の追加検索キーを得示してもよい。

100531

[数3]

$$E(W_i) = M \cdot \left(-\sum_{j=1}^{M} P_{i,j} \log_M P_{i,j}\right)$$

文書 D.中における W.の外理頻度 文書 D.中の追加検索キー候補数

これは、検索結果文書に満瀬なく出現する語は検索キー との翻連が深く。翻連する語を検索条件に加えることで 関連する文書を検索できるという考えに基づいている。

【0054】(2)ユーザが再検索意図を指定すること により、上記第1実験形態と上記変形例(1)の評価値 計算法を切り替えることもできる。

【0055】(3)上記第1実験形線と上記変形例 (1)の両方の評価値計算法による追加検索キーを再検 素素図的に提示することもできる。

【0056】(4)上記第1実施形態では、ある程度の 40 頻度で出現する語を有効語や追加検索キーにするため に 評価値に出現額度を掛けているが、出現頻度を引数

とした単調増加関数の値を掛けても良い。

【0057】(5)上記第1実練形態では、各語の出現 確率のエントロピーを1から減じた値を局在度としてい るか、これに限定されるものではなく、他の演算により

場在度を求めてもよい。 【0058】(6)上記第1実練形態では、ステップS 303でインデックス作成を行ない。その後に適加検索 キーのための処理を行なうよう説明しているが、これに 1.9

デッタス作成を行会うようにしてもよい。

【0059】(7)上記第1実籍形態では、必ず追加検 索キーを提示するよう説明しているが、これに限定され るものではなく、追加検索キーを提示するかしないかを ユーザが特定するようにしてもよい。

【0060】(8)上記算1契純形態では、形態素解析 により有効語候補を推出するよう説明しているが、これ に限定されるものではなく、文字機の違いによって切り 出した文字列を有効器を補としてもよい。

【0061】(9)上記第1実権形態では、名詞を有効 10 るメモリに書き込まれた後、そのプログラムコードの指 語候補として補助するよう説明しているが、これに限定 されるものではなく、他の温詞の文字列や、投合語をも 有効器解稿として抽出してもよい。 い、その処理によって新述した実態が態の機能が実現さ い、その処理によって新述した実態が態の機能が実現さ

【0062】(10)上記算1乗締那線においては、被除案テキスト候棒部101、有効部候補乗度持部105、有効部候補大と「保持部106、有効部と対して、保持部108、検索キー保棒部111、検索結果保持部113、通加検索キー候棒部1175、通加検索キー保持部1175、通加検索キー保持部117を入って、インデークス保持部103、有効連線度保持部110をディスク装置で実現する場合について説明したが、これに限定されるものではな

て、任憲の記憶媒体を用いて実現してもよい。 【0063】(11)上記第1実施彩態においては、各部を同一の計算機上で構成する場合について説明した。

が、これに限定されるものではなく、キートワーク上に 分散した計算機や処理装置などに分かれて各部を構成し てもよい。

【0064】(12)上記第1実験形態においては、ア ログラムをNの時に保持する場合について表明したが、 これに限定されるものではなく。任意の記憶機体を用い 35 て実現してもよい、また、間様の動作をする国際で実現 してもよい。

(10065) (13) 本利明は、複数の機器から構成されるシステムに適用しても、1つの機器からなる装置に 画用してもよい。前途した実施影響の機能を実現するソフトウェアのプログラムコードを記録した記録媒体を、システム或いは装置に接着し、そのシステム或いは装置のコンピュータ(またはで) 日で到回り か記録媒体に 格納されたプログラムコードを記み出し気行することによっても、達成されることは言うまでもない。この場の 記録媒体から認み出せられてプログラムコード自体が 耐張した実施形態の機能を実現することになり、そのフログラムコードを記録した記録媒体は本売明を構成することになり、この一方となる。

【0066】(14) アログラムコードを供給するため の変数解体としては、例えば、フロッピーディスク(登 録簡編)、ハードディスク、光ディスク , 光影気ディス ク、CDーROM、CDーR、磁気テープ、不揮発性の メモリカード、ROMなどを用いることができる。

【0067】(15)コンヒュータが読み出したプログ 切 すブロック間である。

ラムコードを実行することにより。前述した実験形態の 機能が実現されるだけでなく、そのプログラムコードの 指示に基づき、コンピュータ上で舞踊しているらりなど が実際の処理の一部または全部を行ない、その処理によ って前述した実施形態の機能が実現される場合も含まれ ることは言うまでもない。

【0068】 (16) 記録媒体から読み出されたプログ ウムコードが、コンピュータに挿入された機能放張ボー ドやコンピュータに接続された機能放張エニットに構わ るメモリに書き込まれた後、そのプログラスコードの情 がに基づき、その機能放張エニットに 備かるCPUなどが実際の処理の一部または全部を行な い、その処理によって前並した実施が確い機能が実現さ れる場合も会れることは言うまでもなった。

[0069]

【発明の効果】以上詳述したように、第1、第2及び第 4の預用であるテキスト検索装置によれば、検索対象文 書と検索結果文書に適合した再検索を容易に行なうこと 方面像に立る。

3、有效腐額度保持部110をディスク装置で実現する 29 [0070] 第5、第6及び第8の発明であるテキスト 場合について説明したが、これに限定されるものでは今 検索方法によれば、上記第1、第2及び第4の発明と同 ※の効果を参する。

> 【0071】第9、第10及び第12の発明であるコン ビュータ電み取り可能が記録様体によれば、コンビュー 夕によりアログラムを読み出して実行することにより上 記集1 第2万代第42の時1と関係の効果を壊する。

【0072】第3の発明である有効語頻度作成装置によれば、テキストにおける有効語の文書別頻度を簡易かつ 的磁に作成することが可能になる。

(10073)第7の発明である有効語頻度作成方法によれば、上記第3の発明と開発の効果を奏する。

【007-』】第11の発明であるコンピュータ読み取り 可能な記録様体によれば、コンピュータによりアログラ ムを読み出して実行することにより上記第3の発明と問 等の効果を奪する。

【団節の簡単な影明】

【図1】本発明の第1実施形態のテキスト検索装置の機 能構成を示すアロック図である。

【図2】有効語族補リストの例を示す図である。

40 【図3】有効語リストの例を示す図である。

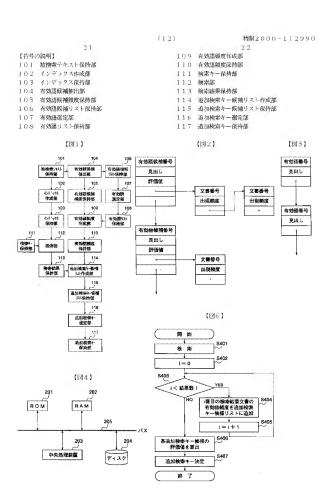
【図4】実施形態に係るテキスト検索装置のハードウェ ア構成を示す図である。

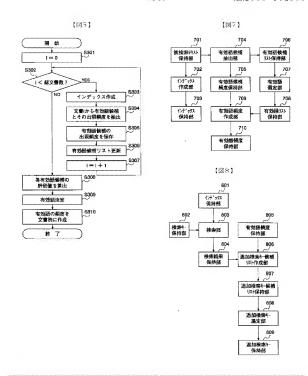
【図5】第1実施形態におけるインデックス作成フェー ズの概要を示すフローチャートである。

【図6】第1実施形態における検索フェーズの概要を示すフローチャートである。

【図7】本発明の第2実施形態の基本的な機能構成を示すプロック図である。

【図8】本発明の第3実施形態の基本的交機能構成を示 すでロック図である





フロントベージの絞ぎ

(73) 売明者 上田 隆也 東京都大田区下丸子 3 丁目 90番 2 号 キヤ ノン株式会社内 (72)発明者 池田 特治 東京都大田区下丸子3丁目30番2号 キヤ ノン株式会社内 ビターム(参考) 59075 MiO3 MiO2 Mid %31 Pt12 Pt22 Pt27 Pt35 P504